融合选择性注意衰减模型的信息简报自动生成方法研究：以 Unesco 科技报告为例[1]

田文波 1 ，宋培彦 1*，吴柯莹 1,2，冯超慧 1
1 天津师范大学管理学院　2 吉林大学商学与管理学院

摘要：[目的/意义]简报是重要的情报产品，摘要和主题词汇集是简报的重要组成部分。联合国教科文组织 Unesco 发布了大量高价值的科技报告，为了满足用户对国际专业知识的需求，需要快速形成信息简报、提高情报服务能力。[方法/过程]本文以认知科学中的“选择性注意衰减”理论模型为基础，将信息简报的生成作为人类认知信息加工的模拟过程，对实现自动生成摘要和主题词集方法进行了探究。首先以选择性注意中的“衰减器”模型为理论支撑，从摘要、主题词、简报三个层级进行一体化设计。然后，采用 KeyBERT 和 Transformer 算法，对联合国教科文组织 Unesco 科技报告进行主题词抽取和摘要生成，形成可参考性的简报情报产品，接着采用信息熵和 ROUGE 值对生成结果进行评价。[结果/结论]实验结果通过 ROUGE-2、ROUGE-L 值评价后表明选择性注意衰减模型能够提高摘要效果，覆盖文本的核心信息；从信息熵的角度进一步论证表明该方法自动生成摘要结果符合人的基本认知水平。研究还发现,将认知科学与计算模型紧密结合对于提高信息简报的可解释性和科学性有显著作用，有助于形成可计算、可解释的信息简报生成与知识服务模式。

关键词：知识发现；选择性注意；文本摘要；主题抽取
分类号：G353

1. 引言

信息爆炸引发了信息过载，用户获取信息的效率面临挑战。通过对海量信息进行自动摘要，并以专题简报形式推送给有关机构或用户，有利于帮助用户把握前沿话题、及时掌握动态、实现科学决策。国际组织所发布的各类报告具有较高的权威性和参考价值，有必要对其进行快速监测和动态跟踪，生成高质量信息简报并推送给有关机构，具有重要应用价值。

从认知科学来看，信息简报本质上是人类对信息再加工的过程。人在阅读文本时，大

---

[1] **基金项目：**本文为全国科技名词委 2020 年科研项目“国际组织术语库集成方法研究”（项目编号：YB20200011）研究成果之一。
作者简介：田文波（ORCID:0000-0002-4251-2763，邮箱：1207159394@qq..com ），硕士研究生，研究方向：知识组织与知识聚合；宋培彦（ORCID:0000-0003-1055-2717，通讯作者，邮箱：spyer2008@126.com ），男，博士、副教授、硕士生导师，研究方向：知识组织、自然语言处理、科技大数据；吴柯莹（ORCID:0000-0003-3013-6657，邮箱:wkeying129@163.com），硕士研究生，研究方向：情报学；冯超慧（ORCID:0000-0001-6503-4111），硕士研究生，研究方向：知识组织与知识服务。

脑往往是特定时间特定环境下获取有限的关键信息，因此通常会将有限的处理能力分配在焦点信息上，即选择性注意机制。人类能够合理分配和充分利用有限的注意力资源，从大

量信息中快速而且精准筛选出高价值信息、过滤低价值信息，以“省力原则”达到高效的信息处理，这是人类在长期进化中形成的一种生存机制。认知科学界提出了“过滤器模型”等理论模型，并通过眼动实验等进行验证，也进一步证实了认知理论模型对实现摘要技术具有有重要理论价值。因此，本文以信息简报生成作为应用场景，以联合国教科文组织 Unesco 数据为例，将选择性注意模型为基础与计算算法紧密结合，模拟人类信息加工内在规律，提出并验证其在信息简报生成中的作用，为提高自动摘要的科学性和可解释性提供有力支持。

2. 相关研究

2.1 注意力模型

注意力与深度学习技术近年来深度融合，发展迅速。曾子明等通过构建基于用户注意力机制的 U-BiLSTM 情感分析模型分析情感演化过程，具有较强的可解释性和准确性，使得 F1 值和准确率都有所提高[1]。周瑛等结合了长短时记忆（LSTM）和注意力机制模型，通过华为 p10 闪存门事件为例，证明了基于选择性注意力机制的情感分析能够提高情感分类的成功率，能够准确的提取情感特征，在短文本的分析中有很好的应用[2]；胡吉明等利用 Text-rank 算法和 CNN-BiLSTM-Attention 集成模型对政策文本进行分类处理，提升了分类的效率和准确度[3]。该领域研究方向主要侧重于将注意力机制与情感分析、文本分类、舆情监测等相结合，对本文文本摘要也有启发。特别是，Transformer 在运算效率和并行处理上有着一定的优势，为实现自动摘要提供了良好的技术条件和方法。值得注意的新动向是，认知模型与计算方法的结合，往往比单纯算法优化和参数设计带来更为可观的科学进步，今后还应该对认知科学加强实质性的结合，形成更多原创性的研究成果。

2.2 文本摘要

现有研究主要是采用机器学习算法，计算机领域和情报技术领域有不少成果。李维等提出了一种将 Text-rank 算法与词向量结合的藏文抽取式摘要生成方法，把句子中每个词语映射到高维词库形成句向量进行迭代，对句子进行评价，选取评价较高的句子作为摘要，从而有效的提升了摘要质量；章成志等设计了基于细粒度评论挖掘的书评摘要方法，为图书信息提供了多维度，细粒度的评价[4]；王晓宇对传统的基于图（graph-based）方法中的文本图构建和词加权方式进行改进，使算法根据句子单词的依存关系，生成多种属性构成语义图，并在此基础上提出融合关键词位置信息、概念层级和连接强度等词权重计算方法，进行重要性

排序，选择高分节点作为关键词集合，有了一定的提高[5]。从用户角度来看，现有摘要的语义连贯性和可读性还有一定的局限，从认知计算角度把不同认知负载的词义和句子向量化，生成以主题为导向的新句子，有望形成更具认知解释力和准确性的新技术。

2.3 主题提取

主题是信息简报的必要组成部分。目前国内主题抽取主要采用以隐狄利克雷分布（Latent Dirichlet Allocation, LDA）算法为代表的主题模型，用无监督学习的方式对全文本进行语义结构和聚类分析，从文本中抽取有价值的主题及主题关键词分布。例如石晶等在LDA为语料库和文本建模的基础上进行文本分析，其结果要明显高于其他方法[6]；曲靖野等运用 LDA 主题模型梳理近 22 年来国内信息服务研究主题演化情况，为该领域的可持续发展提供借鉴和指导[7]。常见的基于语言模型的深度学习方法有 NNLM、word2vec、Elmo、GPT、BERT 等，其中 BERT 是深层双向 Transformer 预训练语言模型，是 NNLM、Word2vec、ELMO 和 GPT 等 embedding 技术的集大成者。李松繁等提出了一种基于 BERT 的前沿研究主题识别方法，实现农业领域前沿研究主题的识别[8]。如何将主题抽取与摘要生成进行一体化设计，还需要进一步研究。

总体而言，学术界对简报生成研究已经有丰富成果，同时还存在两个突出问题，一是技术算法研究多而认知方面的研究偏少，对摘要深层次的理论解释力稍显不足；二是摘要与关键词抽取往往分割，应该对其内在关系进行一体化处理，从而提高简报生成效率和一致性。因此，本文引入了选择性注意“衰减”模型，并与 KeyBert、Transformer 算法结合，为信息简报提出了可行的认知计算依据和实现方案，最后从信息熵和 ROUGE 值两个方面进行量化测评和验证。

3. 理论依据和总体框架设计

3.1 选择性注意衰减模型与认知负载因素

信息简报生成机制可以从认知科学的“注意”研究中得到有益的启示。“注意”是一种聚焦于特定刺激的能力。大多数情况下，注意的聚焦特性是与“选择性注意”相关联的，即人类能够将注意聚焦或分配在一个特定的位置、课题或者信息上。认知心理学家 Broadbent 提出的“过滤器理论”（Filter Theory）认为，人类可以将一些信息过滤掉，从而允许另一些信息得到更高层次的加工，从而免受“信息过载”的困扰。Treisman 提出的“衰减器”理论是对过滤器模型的改进，未被注意的信息并未被真正过滤，而只是被加工的可能有差异而已，进而通过认知实验表明，物理特性、语言、语义都可以用户注意到，并用来吸收信息。其模型如图 1 所示。通过衰减器，各类信息都有机会被进行加工或者过滤，最终形成记忆。认知

负载是指在执行某项认知任务时所需要的认知资源数量。对于熟悉或者简单任务而言，认知负载较低；对陌生或困难任务则需要分配更多的认知资源，其认知复杂相对较高。

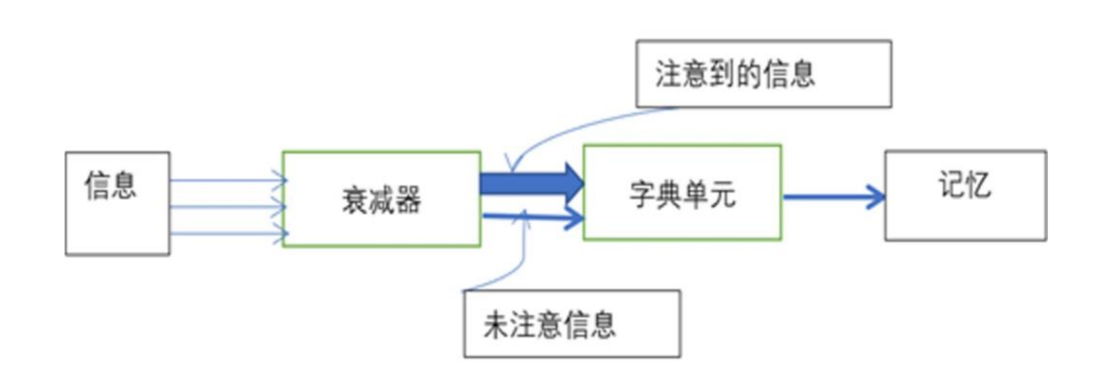


图 1 选择性注意衰减模型理论图

选择性注意衰减模型为信息简报的自动生成提供了重要理论依据。表现在 3 个方面，（1）摘要过程实际上是人人类分配认知资源对信息进行过滤的过程，语言特征和语义是摘要的核心依据，传统依据词频、位置等外在特征的摘要抽取之外，还应该以语义为中心进行抽取。（2）作为语义载体的词汇和句子对人的认知负载有差异，对信息简报生成的作用应该综合考虑。关键词多数属于低负载，而摘要则属于高认知负载，关键词对于摘要属于相容侧干扰项，有助于以较低的认知复杂实现信息的高层加工和特征整合。（3）信息摘要往往采用多重任务并行执行方式生成多个并行的摘要文本，但从非注意任务中获取信息仍有可能，分配注意的能力可以通过信息量或者与语义一致性进行训练，逐步实现自动化加工。本文设计的框架图也是在这一理论的指导下进行，并采用算法进行实现和验证。

3.2 文本自动摘要框架流程图

结合 3.1 节选择性注意力衰减模型理论，本文设计了信息简报生成主要框架，如图 2 所示。

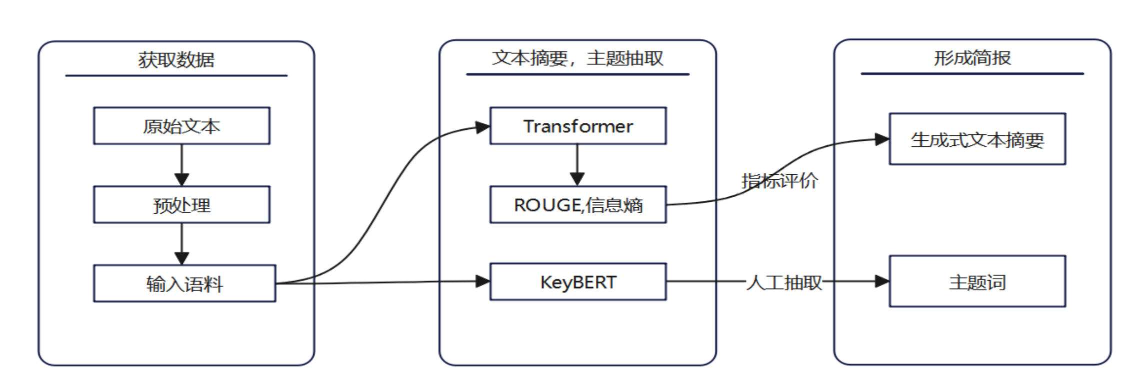


图 2 自动摘要框架流程

该框架包括三个模块，分别是获取数据、文本摘要和主题抽取、生成简报

(1)文本抽取：文本抽取是生成简报的准备条件，首先根据需求选择信息来源，确定目的文本，剔除图表和公式等信息元素。其次导入 jieba 中文分词库对文本进行预处理，去除语气词，标点符号等。

(2)文本摘要、主题爬取： 进行计算之前通过导入 word2ve 库对输入语料向量化处理，再通过 Transformer 计算生成新的语料后，利用 ROUGE 值和信息熵进行指标评价，最后进一步处理形成文本摘要；将语料向量化处理后，调用 KeyBERT 算法库进行关键词抽取形成关键词集，再经过过滤和排序处理后形成主题词，文本摘要和主题词构成了简报的主要内容。

3.3 基于选择性注意衰减模型的计算方法

该理论强调了计算机在文本计算过程中，能够对输入文本内容的重要性进行主动筛选，将更多的注意力分配给更重要的文本上，筛选过滤掉文本中具有干扰性的低负载信息。因此本文以该理论为依据，借助开源算法设计了简报生成的关键方法。

3.3.1 基于 KeyBERT 方法的关键词抽取

主题词是信息简报的基本组成，它是从关键词中抽取出来的能够显示文本结构和语义的词语，可以帮助用户抓住文本重要信息点，实现对文本的主要内容进行整体把握，充当读者阅读摘要的线索。KeyBERT 是一种小型且易用的关键词抽取技术，该算法的核心依然基于选择性衰减注意力，通过对文本进行向量化后，再以语义计算为过程导向进行权值计算，它使用 BERT 嵌入和简单的余弦相似度创建与文档最贴合的关键词或短语。

在本文中，首先利用 BERT 计算文档的 embedding 值，从而获取文档向量级别的表示。然后针对 n-gram 提取词向量，最后利用余弦相似度来确定与文档最相似的关键字或关键短语，就能得到最能描述整篇文档的关键词。在关键词的基础上，结合原文主题内容对关键词进一步筛选，剔除关键词中表达主题性弱的词汇，筛选出的结果能很好的表达文本主题，与生成式摘要结果达成匹配为读者提供多维度的关键信息。

3.3.2 基于 Transformer 模型的摘要句生成

Transformer 基本原理是在 encoder-decoder 的模型中加入了一个加权计算来表示每次词句的影响权重。它通过文本本身的注意力训练从而构建句子之间的关系表示[13]。Transformer 每一层都包含一个多头注意力机制和前馈神经网络，如图 3 所示。

在本文中，先将需要输入的文本进行词向量转换，再通过 encoder-decorder 层进行 attention scores 计算，构建其句子的权重关系，最后经 softmax 计算后输出摘要结果。相比于以往研究抽取式文本摘要或者基于 LSTM 等深度学习摘要，Transformer 通过循环多次的自注意力的计算，不但使得其并行计算效率增加，也提高了其计算性能和计算质量。

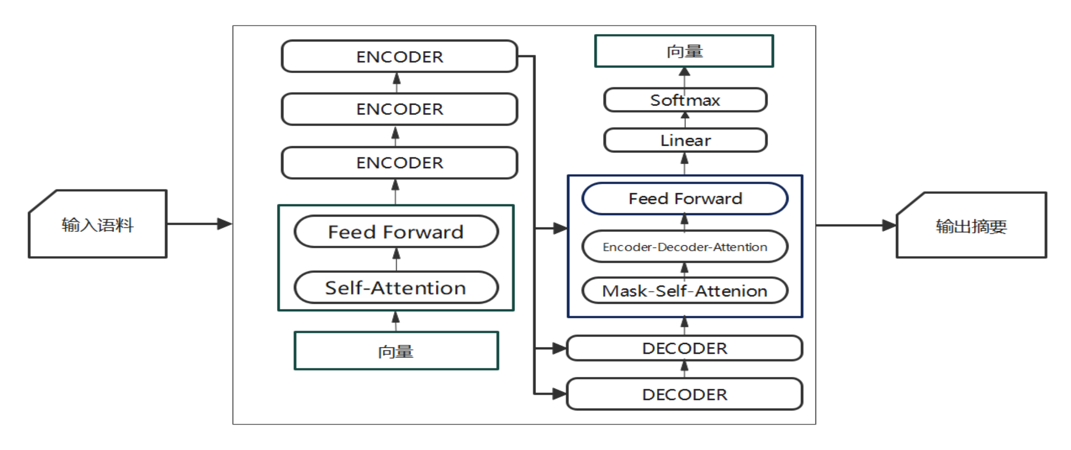


图 3 Transformer 模型图

3.4 简报质量评价指标

3.4.1 简报信息衡量：信息熵

信息熵是用来衡量系统中信息量大小，从文本信息的角度而言，一段文字的熵值越高，则表明这段文本所含的信息量越高。它的应用比较广泛，可以用于验证不同语种表达相同含义下所需文本容量，也可验证特定文本信息含量大小。信息熵的数学公式可以表示为[10]：

$$H(x)=\sum_{x\in X}P(x)log(\frac{1}{P(x)})=-\sum_{x\in X}P(x)log(P(x))$$

其中，x 表示随机变量，它的取值为（x1,x2……xn）,p(xi)表示事件 xi 发生的概率，通常情况下信息发生的概率越大,则这个事件所包含的信息越小，∑p(xi)=1，即所有随机事件的概率和为 1，引入到文本信息量计算中则表示一段文本所包含的随机信息的概率。考虑到文本内容包含语义计算涉及词汇较多，故选取二元模型（bi-gram）对文本进行信息熵的计算[11]。利用信息熵来判断文本摘要信息含量，可从信息量的维度对生成式文本摘要质量进行量化描述。本实验利用 python 语言提供的开源代码构建信息熵计算流程，对生成的摘要进行分词处理后再用 wordtovec 方法进行文本向量化表示，最后通过信息熵公式进行计算并输出结果。

通过对有关信息熵文献和资料的调研可知，不同语言种类文字以及表达方法与信息熵大小有关联如汉语、英语、法语、俄语等信息熵存在差异；同一语言下，不同认知的人群如学者、儿童等描述同一内容的文字信息熵也存在差异。因此通过计算摘要结果的信息熵值，并将该值与中文文本的信息熵标准指标进行对比，可验证该生成式摘要结果的认知水平是否达到标准。

3.4.2 简报质量评测：ROUGE 值

ROUGE（Recall-Oriented Understudy for Gisting Evaluation）是评价文本摘要质量的常用

指标。其通过参考文与实验文本之间共现的词语或句子个数，来判断摘要结果的质量，因此计算该指标需要获得由专家手动生成人工摘要作为标准摘要集，与机器生成的摘要作为对比。用公式表示为：

$$\text{ROUGE - N} = \frac{\sum_{S\in\{ReferenceSummaries\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S\in\{ReferenceSummaries\}} \sum_{gram_N \in S} Count(gram_N)}$$

ROUGE-N，其中 N 表示 N-gram，代表了分母为参考文本中 N-gram 个数，分子为摘要结果中 N-gram 的个数，进行评价时通常以 ROUGE-1，ROUGE-2 为指标。

除以上两个指标外，为增强评价的说服性，本文还参考了最长公共子序列计算方式，即 ROUGE-L，其计算公式为：

$$R_{LCS} = \frac{LCS(C,S)}{len(S)}$$

$$P_{LCS} = \frac{LCS(C,S)}{len(C)}$$

$$F_{LCS} = \frac{(1+\beta^2)R_{LCS}P_{LCS}}{R_{LCS}+\beta^2 P_{LCS}}$$

其中，C，S 分别表示参考文本和实验文本 LCS(C,S),表示为文本 C 与 S 之间最长公共子序列长度，len(S)和 len(C)分别表示两个文本的长度，Rlcs 表示召回率，Plcs 表示准确率，Flcs 即是 ROUGE-L。β 则表示一个非常大的数，因此经过推导公式后发现，Flcs 几乎等 Rlcs。本文生成的摘要主要为短文本，单文档摘要，故采用 ROUGE-2,ROUGE-L 两个值作为评价指标。

4. 实证研究

本文从国际教科文组织网站中选取了 50 篇有关教育话题的新闻或报告作为实验文本；其次利用开源工具对目的文本进行爬取、储存在数据库中，并通过人工方式进行简单清洗和预处理，剔除语气词、连接词、图表、复杂的公式和数字等等，并将文本进行向量化处理，以便于后续程序的进一步计算；最后将文本数据在构建好的模型下进行循环计算，计算出生成式的文本摘要结果。

4. 1 数据采集

本研究数据来自国际教科文组织（UNESCO），选择教育方面的报告作为实验文本，具有可读性和参考价值。按照 3.1 流程图所示，利用 python 中 selemium 库选取了 UNESCO 中 50 个具有代表性的报告进行初步人工预处理后储存在数据库中。其次，对数据库中的文本

数据进行分词，去除停用词等操作后完成预处理。

4.2 文本摘要自动生成

本节按照 3.2 节提出的总体框架进行实例验证。首先，将预处理后的文本导入 python，通过 jieba 分词库进行分词处理，再利用 word2vec 转化成向量表示便于计算。然后，依据选择性注意衰减模型理论，利用开源工具搭建的 Transformer[13]及 KeyBERT 算法框架，将采集数据预处理为机器可识别的词向量后，通过已有框架进行计算，分别生成文本摘要和关键词集。在此基础上，本文建立了可量化的信息熵和 ROUGE 值对文本摘要结果进行量化评价；在关键词集的基础上，抽取围绕着文本主题叙述的关键词作为主题词，并按照文本顺序对关键词进行排序，最终形成规范化的信息简报。部分实验结果如表 1 所示。

表 1 部分生成式文本摘要

序号	结果
文本 1：摘要	联合国教科文组织发起了一项全球范围的 "快乐学校倡议"。学校应该成为支持社会凝聚力的场所，创造跨越差异的社区。学校也应该通过快乐和参与来培养学生对学习的终生热爱，而不是通过将学习成绩置于一切之上来阻止学生学习，从而损害个人的幸福。幸福学校 "倡议正在走向全球，教科文组织正在向全世界发起。它的目的是帮助全球的学生和教育工作者，由联合国教科文组织发起，旨在帮助儿童和教育工作者。它将成为全世界学校和教育工作者的一个全球模式，这是一个国际学校教育的国际模式，以帮助整个大陆和国家的年轻人和教育工作者，它是由亚洲教育的全球教育研究所和亚洲国际教育理事会以及世界教育理事会的欧洲理事会发起的。
文本 1：主题	联合国教科文组织，全世界，年轻人，幸福危机，生产力，压力，难民，学校，提高，学生，凝聚力，平等。
文本 2：摘要	2017 年，在全球范围内，超过三分之二的国家中，学习工程、制造和建筑或信息和通信技术（ICT）的人中只有不到四分之一是女性。在几乎所有的教育系统中（87%），男孩比女孩更经常回答他们想从事涉及数学的工作。男孩和女孩在数学领域工作的愿望与他们对自己在该学科的能力的信心密切相关。这表明，解决女孩对科学和数学的信心问题应该继续成为政策制定者的关注点。国际教育协会指南针指出，这可能会导致更少的表现出色的女孩进入 STEM 高等教育领域教育系列简报。该报告发表在 IEA 的 TIMSS 2019 年数据特刊上，样本为 25 万名学生，显示 8 年级时希望从事数学或科学相关职业的男孩多于女孩。
文本 2：主题	联合国教科文组织，国家，科学研究，统计学，四分之一，女孩，取得成功，成就，分析，信心，不同
文本 3：摘要	在东南亚，如果一个孩子的温度高于平均水平 2 个标准差，预计他将减少 1.5 年的学业。学年平均温度上升 1.8°C，学习成绩就会下降 1%，超过 32.2°C的 6 天也是如此。假设最佳温度低于 22°C，教室里的温度从 30°C 降到 20°C 将使考试成绩平均提高 20%。污染的空气会大大降低认知能力，也许是不可逆转的，因为怀疑有神经毒性。在以色列，即使暂时暴露在灰尘中，也会降低学生的考试成绩和中学后的教育程度。一项对西班

	牙巴塞罗那近 3000 名儿童的队列研究发现，在调整了社会经济地位后，那些暴露在西班牙高污染水平下的儿童，其认知发展的增长低于污染较少的学校的同龄人的发展增长。
文本 3：主题	阿拉伯联合酋长国，高温，高污染，平均气温，造成，考试成绩，中学，年限，降低，认知，发展

基于选择性注意衰减理论的计算框架，将文本摘要与主题抽取深入到了语义理解层次，再重新组织文本生成新的语料[14]，使得摘要结果自然，具有较强解释力，相对比抽取式摘要文本衔接更为流畅。

通过将结果进一步处理，计算文本摘要的信息熵，对 10 篇文献的摘要信息含量进行评价，计算结果如图 4 所示，横坐标轴代表了十个计算结果，纵坐标代表了信息熵，柱状图顶部显示了每个摘要结果的信息熵值。

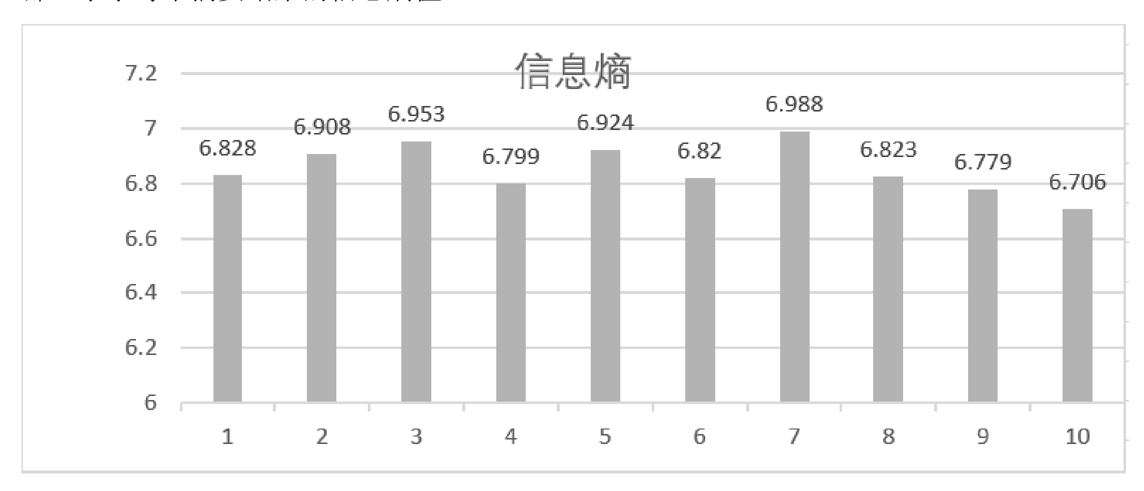


图 4 摘要质量测评：信息熵

通过随机对十个摘要结果进行了熵值计算并统计，对表 1 的文本摘要结果利用信息熵计算可得所抽取样本信息熵平均值达到 6.852；通过文献和资料调研对中文文本信息熵实验结果调研发现，基于 2-gram 的中文文本信息熵范围介于 4-8 之间，表明该方法生成的摘要结果信息熵达到中文文本信息熵标准，符合人的认知水平。

除此以外，本实验从 ROUGE 值进一步验证摘要结果的质量。验证召回率需要合适的参考摘要，为了使研究结论更具有客观性，本文采取人工方法分别对所抽取的十个文本生成摘要作为参考答案，便于计算召回率。计算结果如下图 5 所示，参考 3.4.2 的信息熵公式可知该折线图显示的结果是通过人工生成参考摘要与实验得到摘要之间的词共现率实现的，从折线图的趋势可以看出，由于 ROUGE-2 运算时考虑连续两个词向量，而 ROUG-L 是依赖连续多个词向量, ROUGE-2 值普遍大于 ROUGE-L 值。

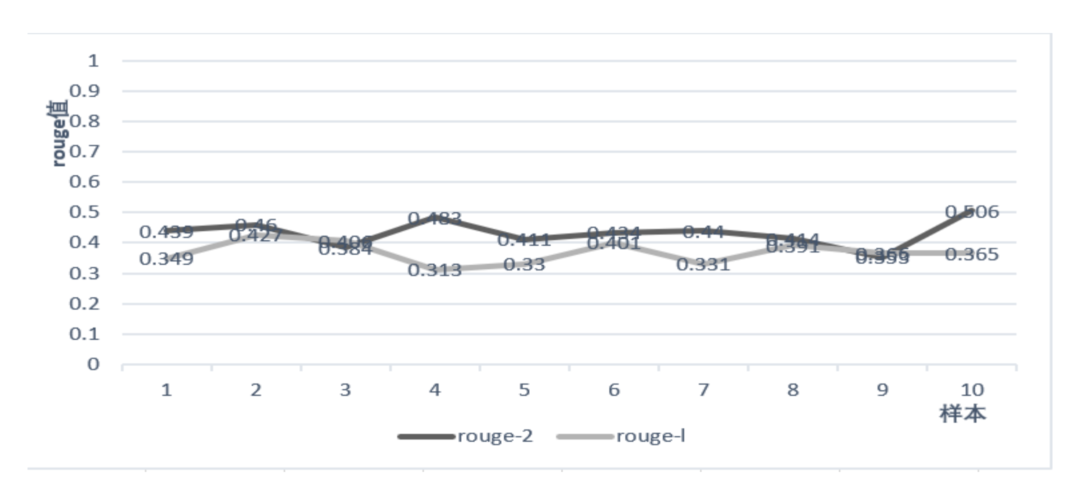


图 5 摘要质量测评：ROUGE 值

选择 ROUGE-2，ROUGE-L 参考标准作为指标进行评价，评价的标准为生成式文本摘要与参考人工摘要的文本复现率。本文所抽取样本在 ROUGE-2、ROUGE-L 两个指标平均值分别为 0.432 和 0.367，该结果与文献《基于最大边界相关度的抽取式文本摘要模型研究》等多篇文章对基本持平[12]，但生成式摘要的可读性更高。究其原因，选择注意衰减理论为指导的生成式文本摘要本质上是通过深层次的语义计算实现对高认知负载信息加工效果，并对这些文本进行重新组织生成简洁、自然的语句。

4.3 简报生成

简报包括主题词和摘要两部分。主题词的作用是帮助读者快速了解文章的关键信息点。主题词所提供的信息较为发散、属于低认知负载的信息加工，读者可以通过主题词了解文章的主要脉络。摘要是文章信息的浓缩，包含了文章的具体信息，读者可以阅读摘要来理解文章大意[15]。通过摘要生成和主题抽取两个维度的信息处理，能够较为充分描述文本信息。

主题词与文本摘要作为简报产品的重要组成成分，相辅相成，在内容和主题把握上有着较强的联系：摘要是主题词的扩展深化部分，主题词是摘要的核心信息。文本摘要和主题词并行计算、相互参照，通过若干主题词的描述可以一定程度上对文本要点进行挖掘，共同构成信息简报。

5. 结论

本文将认知科学中的选择注意力衰减理论和信息简报生成方法相互结合，设计了融合关键词抽取和生成式摘要的信息简报生成框架，以 Unesco 数据为例，采用 Transfomer 和 KeyBERT 方法进行了实证研究，实验结果表明，选择性注意衰减模型能够兼顾关键词、句子和篇章，符合信息处理的认知负载水平，具有较强的解释力和科学性，而且在技术方面则采用信息熵和 ROUGE 值进行评价，根据不同认知负载水平生成不同颗粒度的信息简报，对

于实现情报监测、决策支持有一定的应用价值。今后，对选择性注意的影响因素和机制进一步细分和探究，有助于将认知科学与情报技术有机结合，提高情报产品可解释性与准确性，是值得探讨的方向。

参考文献：

[1] 曾子明,孙晶晶.基于用户注意力的突发公共卫生事件舆情情感演化研究——以新冠肺炎疫情为例[J].情报科学,2021,39(09):11-17.

[2] 周瑛,刘越,蔡俊.基于注意力机制的微博情感分析[J].情报理论与实践,2018,41(03):89-94.

[3] 胡吉明,付文麟,钱玮,田沛霖.融合主题模型和注意力机制的政策文本分类模型[J].情报理论与实践,2021,44(07):159-165.

[4] 章成志,童甜甜,周清清.基于细粒度评论挖掘的书评自动摘要研究[J].情报学报,2021,40(02):163-172.

[5] 王晓宇,王芳.基于语义文本图的论文摘要关键词抽取算法[J].情报学报,2021,40(08):854-868.

[6] 石晶,范猛,李万龙.基于 LDA 模型的主题分析[J].自动化学报,2009,35(12):1586-1592.

[7] 曲靖野,陈震,胡轶楠.共词分析与 LDA 模型分析在文本主题挖掘中的比较研究[J].情报科学,2018,36(02):18-23.

[8] 李松繁,黄永,杨金庆.基于 BERT 的农业领域前沿研究主题识别方法研究[J].情报工程,2021,7(05):100-114.

[9] 刘啸剑,谢飞,吴信东.基于图和 LDA 主题模型的关键词抽取算法[J].情报学报,2016,35(06):664-672.

[10]马峥.基于“反事实”思想测度学术期刊对知识系统信息熵变化的贡献[J].情报学报,2022,41(07):745-761.

[11]程慧平,程玉清.基于 AHP 与信息熵的个人云存储安全风险评估[J].情报科学,2018,36(07):145-151.

[12]余传明,郭亚静,朱星宇,安璐.基于最大边界相关度的抽取式文本摘要模型研究[J].情报科学,2021,39(02):34-43.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EOL].[2022-08-01] https://dl.acm.org/doi/10.5555/3295222.3295349.

[14] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[15] Liu Y, Lapata M. Text summarization with pretrained encoders[J]. arXiv:1908.08345, 2019.

作者贡献说明（独立作者不需提供）

每个作者对成果的贡献应在本部分以单独行的方式进行简要描述。

格式：

作者 1：田文波，提出研究思路，研究方案设计，论文写作与修改

作者 2：吴柯莹，实验设计与实现，论文修改完善

作者 3：冯超慧，实验数据分析，论文修改完善

A study on Automatic Information Briefing Generation Method based on Selective Attention Decay Model: the example of Unesco*

Tian Wenbo[1],*Song Peiyan*[1*],*Wu Keying*[1,2],*Feng Chaohui*[1]

1. *School of Management, Tianjin Normal University,Tianjin,300382,China*

2. *School of business and management, Jilin University, 130015,China*

Abstract: ［Purpose /significance］Briefings are an important intelligence product. Unesco publishes a large amount of high-value specialized literature, and in order to meet the needs of users for international expertise, it is necessary to quickly develop information briefings and improve intelligence services. [Method /Process] Based on the "Selective Attention Decay Theory" in Cognitive Science, this paper investigates the Generation of Information Briefings as an analog process of human cognitive information processing, and investigates the method of Automatic Cross-Linguistic Summary Generation. Based on the " Decay Model" of Selective Attention, an integrated design was developed using the Cognitive Load Capacity at three levels: topics, topic sentences, and briefings. Then, the KeyBERT and Transformer algorithms were used to extract topic words and generate abstracts for the scientific and technical reports published by Unesco, and the information briefings were generated quickly and evaluated using Information Entropy and ROUGE Values. [Result/conclusion] The experiments show that the method has advantages in Information Entropy and ROUGE-2 and ROUGE-L Values, indicating that the Selective Attention Decay Model can improve the summary effect and cover the core information of the text. The study further finds that the close integration of Cognitive Science and Computational Models has a significant effect on improving the interpretability and scientific quality of information briefings, resulting in a computable and interpretable model for Information Briefing Generation and knowledge services.

Keywords: Knowledge Discovery; Elective Attention; text summarization; topics extraction